AEM
Journals.ASM.org

Genotypic and Phenotypic Analysis of Dairy *Lactococcus lactis* Biodiversity in Milk: Volatile Organic Compounds as Discriminating Markers

Amandine Dhaisne,[a,b,c] Maeva Guellerin,[a,d] Valérie Laroute,[a,b,c] Sandrine Laguerre,[a,b,c] Muriel Cocaign-Bousquet,[a,b,c] Pascal Le Bourgeois,[a,d] Pascal Loubiere[a,b,c]

Université de Toulouse; INSA, UPS, INP; LISBP, Toulouse, France[a]; INRA, UMR792 Ingénierie des systèmes biologiques et des procédés, Toulouse, France[b]; CNRS, UMR5504, Toulouse, France[c]; Laboratoire de Microbiologie et Génétique Moléculaires, UMR5100, CNRS, Université Paul Sabatier, Toulouse, France[d]

The diversity of nine dairy strains of *Lactococcus lactis* subsp. *lactis* in fermented milk was investigated by both genotypic and phenotypic analyses. Pulsed-field gel electrophoresis and multilocus sequence typing were used to establish an integrated genotypic classification. This classification was coherent with discrimination of the *L. lactis* subsp. *lactis* bv. diacetylactis lineage and reflected clonal complex phylogeny and the uniqueness of the genomes of these strains. To assess phenotypic diversity, 82 variables were selected as important dairy features; they included physiological descriptors and the production of metabolites and volatile organic compounds (VOCs). Principal-component analysis (PCA) demonstrated the phenotypic uniqueness of each of these genetically closely related strains, allowing strain discrimination. A method of variable selection was developed to reduce the time-consuming experimentation. We therefore identified 20 variables, all associated with VOCs, as phenotypic markers allowing discrimination between strain groups. These markers are representative of the three metabolic pathways involved in flavor: lipolysis, proteolysis, and glycolysis. Despite great phenotypic diversity, the strains could be divided into four robust phenotypic clusters based on their metabolic orientations. Inclusion of genotypic diversity in addition to phenotypic characters in the classification led to five clusters rather than four being defined. However, genotypic characters make a smaller contribution than phenotypic variables (no genetic distances selected among the most contributory variables). This work proposes an original method for the phenotypic differentiation of closely related strains in milk and may be the first step toward a predictive classification for the manufacture of starters.

The mesophilic lactic acid bacterium *Lactococcus lactis* is one of the most extensively exploited microorganisms; it is used in particular in the manufacture of dairy products. Because of its industrial importance, *L. lactis* has been used as a model bacterium for academic and application-oriented studies. Taxonomically, *L. lactis* has three subspecies, *L. lactis* subsp. *horniae*, *L. lactis* subsp. *lactis* (including *L. lactis* subsp. *lactis* bv. diacetylactis), and *L. lactis* subsp. *cremoris* (1). Only *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris* are used in starter cultures for dairy production. *L. lactis* subsp. *lactis* colonizes a wide ecological niche (dairy products, as well as animals and plant surfaces) (2), whereas *L. lactis* subsp. *cremoris* is found mostly in dairy environments (3, 4). Recent genotyping methods involving DNA fingerprinting analysis, including pulsed-field gel electrophoresis (PFGE) and comparative genomic hybridization, have been evaluated for their suitability for characterizing genomic diversity and identifying reliable genetic markers for phenotypic subspecies differentiation (5–9). Multilocus sequence typing (MLST), at the level of the gene, has proved to be a powerful method for describing *L. lactis* population structure and phylogeny with a limited number of genes sequenced (5, 6). Recent MLST analysis of *L. lactis* subsp. *lactis* (6) confirmed *L. lactis* population adaptation to the milk niche and led to the proposal of a new classification into two ecotypes, one corresponding to "domesticated" strains essentially isolated from dairy starters or fermented products and the other corresponding to "environmental" strains isolated from various sources such as plants, animals, and raw milk. The domesticated strains make only a small contribution to the genetic diversity of *L. lactis* subsp. *lactis*. Indeed, phylogenetic analysis indicates that they essentially form two clonal complexes (CCs) that probably emerged only recently from a single founder event (6).

The characteristics of *L. lactis* subsp. *lactis* used in starter cultures for fermented milk dairy production determine their dairy phenotype, which is related to efficient growth in milk associated with fast coagulation. Coagulation is due to acidification and preserves milk from unwanted bacterial or mold growth. However, *L. lactis* subsp. *lactis* also contributes to the organoleptic quality (texture and sensory) of the fermented products. Strain selection has focused on diverse but specific technical characteristics: acidification (10), redox potential (11), texture (12), and physiological features like aroma production (13, 14), nisin synthesis (15), autolysis (16, 17), and enzymatic activities (18). In these previous studies, selection was restricted to one or a limited number of criteria and the behavior of the strains as a whole cannot be predicted. Phenotypic starter screening requires exhaustive study in a controlled dairy environment to evaluate strain performance, but this is a very complex and time-consuming approach that is incompatible with the screening of strain collections. Genotyping

Received 2 April 2013 **Accepted** 22 May 2013
Published ahead of print 24 May 2013
Address correspondence to Muriel Cocaign-Bousquet, cocaign@insa-toulouse.fr.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.01018-13.

doi:10.1128/AEM.01018-13

TABLE 1 Dairy *L. lactis* subsp. *lactis* strains used in this study

Strain	Origin	CC	ST	Presence of:			Reference
				citP	*lacE*	*prtP*	
S86	Starter	CC1	10	−	+	+	6
UCMA5713	Grassland (France)	CC1	18	−	+	+	6
S87	Starter	CC1	15	+	+	+	6
EIP41A	Raw milk (France)	CC1	15	−	+	+	This study
EIP33A	Raw milk (France)	CC1	15	−	+	+	This study
DIA-A	Starter Choozit DIA A FRO (France-Danisco)	CC1	15	+	+	+	This study
EIP33F	Raw milk (France)	CC1	34	+	+	+	This study
EIP37F	Raw milk (France)	CC1	34	+	+	+	This study
MC70	Starter Choozit MC-70 FRO (France-Danisco)	CC2	7	−	+	+	This study

and genomic sequencing are efficient and have been largely automated. There was an expectation that this approach could be used to classify and discriminate between strains and predict their phenotypes (19). Several studies combining genomic, transcriptomic, and phenotypic data for *L. lactis* have been conducted in recent years (7, 8, 20–23), but they have failed to establish strong exploitable links among the genotype, its expression, and dairy phenotypes. Rather, they have highlighted significant transcriptomic polymorphisms among *L. lactis* subsp. *lactis* strains with the same dairy origin (21) and the difficulties associated with explaining and predicting a phenotype other than the presence or absence of relationships involving the gene (19, 20). Understanding the determinants of the phenotypic diversity of dairy strains is a prerequisite for predictive approaches and requires phenotypic screening in environmental conditions that are as close as possible to those encountered during the process of interest (18).

Therefore, this work reports an integrated approach to the assessment of the phenotypic biodiversity of nine *L. lactis* subsp. *lactis* strains exhibiting specific dairy phenotypes and selected as being representative of the "domesticated" ecotype in milk. Strain phenotypic signatures were compared with strain genotypic diversity, revealing discrepancies between the two classifications and highlighting the need for an integrated genotypic and phenotypic classification that takes into account both aspects.

MATERIALS AND METHODS

Bacterial strains and growth conditions. Nine strains of *L. lactis* subsp. *lactis* were selected from commercial starter cultures (Danisco-France) and from various laboratory collections (Laboratoire de Microbiologie et Génétique Moléculaires, Université Paul Sabatier, Toulouse, France, for Sx and EIPx strains; LMA, Caen, France, for UCMAx strains). The origins and characteristics of the strains used in this study are listed in Table 1. Stock cultures were kept frozen (−20°C) in M17 broth (Oxoid Ltd., Basingstoke, Hampshire, England) containing 2% (wt/vol) lactose and 20% (vol/vol) glycerol. Individual strains were first grown at 37°C in sterilized (110°C, 15 min) reconstituted skim milk (10.9%, wt/vol). These overnight cultures were used to inoculate (10%, vol/vol) 500 ml of pasteurized standardized cow's milk (fat, 35 g/kg, protein, 40 g/kg) in 1-liter flasks. The pH was unregulated (initial value of 6.60). Fermentations were continued at 28°C until the pH reached a value of 4.60, and then the milk was transferred to 4°C. The fermented milk was stored in the dark at 4°C for 1 (sample A) or 14 (sample B) days before analysis. Six biological replicates (independent batches) were performed for each strain.

Growth measurements and acidification activity. Bacterial growth in milk (μ, h^{-1}) and cell cultivability after 1 and 14 days of storage were determined in triplicate by plating appropriate dilutions onto M17 agar, incubating them at 30°C for 48 h, and counting the colonies (CFU · ml^{-1}). Cell cultivability was calculated as the ratio of the viable count after 14 days to that after 1 day.

Acidification activity of lactococcal cells was monitored by using the CINAC system (Alliance Instrument). For each culture, four descriptors were defined to characterize the acidification activity during fermentation: the maximal acidification rate (V_{max}), expressed in 10^{-3} pH units per minute and calculated from the slope of the pH curve as a function of time; the time (in hours) necessary to reach V_{max}; the pH at which V_{max} is attained; and the time (in hours) necessary for the pH to decrease to 4.60.

Fermentation analysis. Substrate (lactose and citrate) and fermentation end product (lactate, formate, acetate, acetoin, diacetyl, and ethanol) concentrations in fermented milk after 14 days of storage were determined by high-pressure liquid chromatography (HPLC) as previously described (24). Briefly, a solution of 0.3 M barium hydroxide and zinc sulfate was used to precipitate proteins from the samples. The samples were then analyzed in a Bio-Rad HPX87H column maintained at a temperature of 48°C in a 1200 series preparative HPLC apparatus (Agilent Technologies, Waldbronn, Germany), and 5 mM H_2SO_4 was used as the eluent at a flow rate of 0.5 ml · min^{-1}. Free amino acid (FAA) concentrations in milk culture supernatants at 14 days were measured by HPLC as follows. Proteins in the samples were precipitated by adding 4 volumes of methanol to 1 volume of the sample and incubating the mixture overnight at 4°C. The mixture was then centrifuged, and the supernatant was kept for amino acid analysis as previously described (25). The amino acids were automatically derived with *ortho*-phthalic aldehyde (OPA) and 9-fluorenylmethyl chloroformate (FMOC), and the derivatives were separated on a Hypersil AA octadecyl silane column (Agilent Technologies, Waldbronn, Germany) at 40°C by using a linear gradient of acetate buffer (pH 7.2) with triethylamine (0.018%), tetrahydrofurane (0.03%), and acetonitrile (60%). A diode array detector was used at 338 nm for OPA derivatives and at 262 nm for FMOC derivatives. These analyses were performed on the six biologically independent batches for each strain.

Identification and semiquantification of volatile organic compounds (VOCs) in fermented milk. VOCs in fermented milk were identified and semiquantified by the Analytical Chemical Service of ISVV (Institut des Sciences du Vin et de la Vigne), Bordeaux, France. Separation and semiquantification were carried out by solid-phase microextraction (SPME)-gas chromatography (GC)-mass spectrometry (MS) analysis.

The internal standards used for GC analysis were [$^2H_5$]ethyl acetate (ethyl acetate-d_5) and [$^2H_{15}$]octanoic acid (octanoic acid-d_{15}) from Sigma-Aldrich (Saint-Quentin-Fallavier, France) and [2,2-^{2}H]3-methylbutanal (3-methylbutanal-d_2) from Euriso-top (Saint-Aubin, France). Solutions were prepared at 100 mg · $liter^{-1}$ with MilliQ water and kept at 4°C.

A 5-g sample of fermented milk was mixed with 250 μl of internal standard mix solution (at 10, 20, and 20 mg · $liter^{-1}$ for ethyl acetate-d_5, octanoic acid-d_{15}, and 3-methylbutanal-d_2, respectively) in a 20-ml glass sample vial. The vial was then tightly sealed, and the contents were homogenized with a vortex shaker at 55°C for 30 min. The SPME fiber

(Supelco, Bellefonte, PA) used was coated with a 50/30-μm layer of divinylbenzene-carboxen-polydimethylsiloxane (57299-U; StableFlex) and maintained in the headspace at 55°C for 90 min. The volatile analytes adsorbed to the SPME fiber were analyzed by GC-MS with an HP 5890 GC system coupled to an HP 5972 quadrupole mass spectrometer (Agilent Technologies, Wilmington, DE) equipped with a Gerstel MPS2 autosampler. The compounds were separated on a BP 21 capillary column (60 m by 0.32 mm, 0.25-μm film thickness; SGE, Courtaboeuf, France). The injection port was programmed to heat at 250°C. The oven temperature was programmed at 40°C for 5 min, raised to 240°C at 3°C · min^{-1}, and then held at that temperature for 10 min. Volatile compounds were identified in SCAN mode (mass range of 29 to 200 *m*/*z* at a scan rate of 1.49 scans · s^{-1}), and the ionization energy was set at 70 eV. Only compounds for which the area was at least 2-fold the background noise were selected. The eluted compounds were identified by their retention times and by comparison of their mass spectra with those in the National Institute of Standards and Technology database (http://www.nist.gov/srd/ [July 2002 version]).

Selected volatile compounds were detected and semiquantified with the mass spectrometer operating in selected-ion-monitoring mode with electron ionization at 70 eV. VOCs were semiquantified by determining the ratio of the total ion count (TIC) of each compound to the TIC of the corresponding internal standard for 1 g of fermented milk. The internal standard was chosen according to the chemical function and the retention time. For each strain, analytical triplicates were performed. Detection and quantification limits and coefficients of variation were determined for each compound.

DNA manipulation and phylogenetic analysis. Bacteria were grown at 30°C on M17 broth. Genomic DNA was extracted from overnight cultures with the DNeasy tissue kit according to the manufacturer's instructions (Qiagen, Hilden, Germany). Genetic markers of important industrial traits (*lacE*, encoding lactose-specific enzyme II of the phosphotransferase [PTS] system; *prtP*, encoding the cell envelope-associated serine proteinase; and *citP*, encoding the membrane-bound citrate permease involved in citrate uptake) were detected by PCR amplification and standard agarose gel electrophoresis. PCR primers for the *lacE* (lacE-F2, 5′-AGCGTCTATGGTAGGGTTCC-3′; lacE-R2, 5′-GATGGCACGGTTACGATCTG-3′; PCR product size of 606 bp), *prtP* (prtP-F2, 5′-GAGGCAGTGAAACTGTTAGTC-3′; prtP-R2, 5′-TCATTCGCAGCAGTACATC-3′; PCR product size of 713 bp), and *citP* (CitP1, 5′-ATGATGAATCACCCG-3′; CitP2, 5′-ACTTCATGAATATGAC-3′; PCR product size of 1,327 bp) genes were designed by standard procedures with Clone Manager version 9.0 software (Sci-Ed Software). The cycling conditions used were 95°C for 5 min followed by 30 cycles of 95°C for 30 s, 55°C for 1 min, and 72°C for 1 min for the *lacE* and *prtP* genes or 98°C for 5 min followed by 30 cycles of 98°C for 10 s, 42°C for 30 s, and 72°C for 1 min for the *citP* gene, using an MJ Mini thermocycler (Bio-Rad, Hercules, CA). Each PCR involved a 25-μl mixture containing 10 ng of genomic DNA, 0.5 mM each primer, and 12.5 μl of iQ SYBR green Supermix (Bio-Rad, Hercules, CA).

MLST analysis was performed by the six-locus MLST scheme exactly as described in reference 6. The concatenated sequences generated were used for phylogenetic analysis with MEGA5 software (26).

For PFGE, preparation of genomic DNA embedded in an agarose matrix, digestion of DNA by restriction endonucleases, and electrophoresis were performed as previously described (27). The genomic relatedness of bacterial strains was estimated from pairwise comparisons of PFGE SmaI macrorestriction patterns, and a matrix of binary data was constructed from the presence or absence of each band. Dice coefficients (S_Ds) and corresponding genomic distances ($1 - S_D$) for each pairwise comparison were calculated from the matrix of binary data with the WINDIST program (28). Dendrograms based on the unweighted-pair group method using average linkages (UPGMA) were constructed with the NEIGHBOR program of the PHYLIP package v3.69 (29).

Data analysis. All variables (physiological descriptors, metabolite concentrations, and genetic distances) were first normalized (centering [subtracting the population mean] and then scaling [dividing this difference by the standard deviation of the population]) to allow an unbiased comparison of these heterogeneous data. Missing data in COV semiquantification for a third of the biological replicates were filled in by using the geometric mean of the available replicates of that strain. To investigate the relationships between strains and the phenotypic variables of fermented milk, an analysis of variance (ANOVA) of the 82 phenotypic variables was performed. The level of significance for all statistical analyses was set to a *P* value cutoff of 0.05. Spearman's rank-order correlations with *P* values adjusted by the Benjamini-Hochberg method to control the false-discovery rate (30) were performed to investigate pairwise associations between the variables involved in strain phenotypes in milk. Multidimensional methods such as principal-component analysis (PCA), hierarchical ascendant classification (HAC), partial least-squares discriminant analysis (PLS-DA), and sparse PCA were carried out with R free statistical software (2.15.0 version) and the mixOmics package (31). The Euclidian distance metric and Ward's criterion were used for HAC in phenotypic and genotypic classifications to constitute hierarchical groups of mutually exclusive subsets in which members are maximally similar with respect to their specified characteristics (32). Bootstrap analysis was used with 1,000 simulations by cluster analyses. Consensus trees were built by bootstrapping (bootstrap value >0.50). To identify discriminatory variables, a variable selection method was developed on the basis of the principal components (PCs). The variable selection combined a test of strain dependency (removal of one strain from the sample in order to differentiate markers highly dependent on a strain from common markers) and a selection of the most contributory variables for this set of individuals. The contribution of a variable to the selected PCs is obtained by the ratio of the sum of the squared factor score divided by the sum of the eigenvalues. The contributory threshold was set to 75% for this procedure; therefore, only variables with a contribution above the third quartile of the contributory distribution were selected as most contributory variables. This procedure was done iteratively for each strain, and strain clustering based on the variables selected was then performed. The cluster robustness and consequently the reliability of the selected variable set were assessed by bootstrapping. In this work, the variables selected represented 25% of the total variables (i.e., 20 variables for phenotypic analyses and 22 for phenotype-genotype investigation, respectively) with the highest head count among each set of the most contributory variables. In order to provide a deeper analysis of the robustness of the variable selection method developed, results were compared to those of variable selection through sparse PCA (33).

RESULTS

Genotypic characterization of *L. lactis* subsp. *lactis* strains. To investigate the phenotypic diversity of *L. lactis* subsp. *lactis* strains under growth conditions close to those of relevant industrial processes, nine strains isolated from raw milk and starter cultures (Table 1) were selected according to their dairy phenotype as pure cultures (ability to grow efficiently in milk inducing fast coagulation). All of these strains were considered to belong to the "domesticated" ecotype (6). The genetic markers of industrial traits checked by PCR amplification were the *lacE* gene, encoding lactose-specific enzyme II of the PTS system, and the *prtP* gene, encoding the cell envelope-associated serine protease. Both were present in all nine strains. *L. lactis* subsp. *lactis* bv. diacetylactis isolates are widely used in the dairy industry because of their ability to take up and consume citrate to produce diacetyl and acetoin, two components essential for a creamy and buttery aroma (8, 18, 22). Four strains in our panel were assigned to *L. lactis* subsp. *lactis* bv. diacetylactis on the basis of results of PCR amplification of *citP*, a plasmid-borne gene encoding the permease involved in citrate uptake (34).

"Domesticated" *L. lactis* subsp. *lactis* currently contains 13 se-

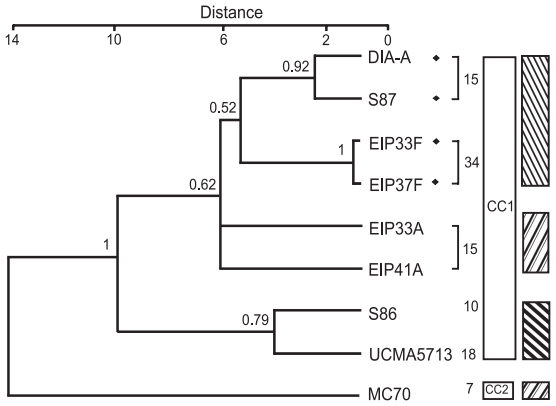

FIG 1 Integrative genotypic classification of nine domesticated *L. lactis* subsp. *lactis* strains from genetic (MLST) and genomic (PFGE) data sets. The Euclidian distance metric and Ward's criterion were used for hierarchical ascendant clustering on scaled MLST and PFGE distances. From left to right, a strain dendrogram, strain names, the presence of the *citP* gene (black diamonds), STs, CCs, and the three main clusters based on PFGE classification (shading) are shown. Bootstrap values above 0.50 are indicated at the corresponding nodes ($n = 1{,}000$).

quence types (STs) organized into two main CCs (https://www-mlst.biotoul.fr/Lactococcuslactissubsplactis/). The main CC (CC1) contains 11 STs (ST1, ST6, ST9, ST10, ST15, ST16, ST18, ST22, ST23, ST34, and ST36), with ST15 predicted to be the ancestor genotype, whereas CC2 comprises ST7 and ST4. MLST analysis revealed that eight of the nine strains used in this study belonged to CC1 (four strains of ST15, two strains of ST34, one strain of ST10, and one strain of ST18) and that one in CC2 was ST7 (Table 1). We therefore considered our strain sample to be genetically representative of the "domesticated" ecotype. As strain redundancy in bacterial collections cannot be excluded, especially for strains belonging to the same ST, SmaI macrorestriction analysis and PFGE were used to confirm that each of these strains was unique and different from the other eight. The genomic relatedness of the selected strains was assessed by computing S_Ds from pairwise comparisons of the SmaI macrorestriction patterns (see Fig. S1A in the supplemental material): 52% of the calculated S_D values were lower than 0.35, the value observed when comparing the two different subspecies *L. lactis* subsp. *cremoris* and *lactis* (6). These findings confirmed previous descriptions of the substantial genomic variability within *L. lactis* subsp. *lactis* (6). The consensus tree built from UPGMA-based clustering of PFGE distances yielded three major clusters (bootstrap value >0.50) and clustered the four strains of *L. lactis* subsp. *lactis* bv. diacetylactis together (bootstrap value, 0.61) (see Fig. S1B). However, strains of the same ST or the same CC were not necessarily grouped together.

We thus considered unsupervised hierarchical clustering integrating both genetic (MLST) and genomic (PFGE) data sets as an alternative approach to improve strain clustering by taking into account genotypic aspects. For this purpose, HAC with the Ward criterion was performed to constitute hierarchical groups of mutually exclusive subsets (32). The accuracy of the genotypic dendrogram was assessed by bootstrapping, and a consensus tree was built (bootstrap value >0.50) (Fig. 1). This integrated genotype classification is consistent with the mean features of gene phylogeny, *L. lactis* subsp. *lactis* bv. diacetylactis discrimination, and strain genomic uniqueness (Fig. 1). However, it failed to group strains of ST15 in one class. The latter result can be explained by evolutionary concepts. Knowing that clonal diversification of lactococcal strains is mostly dependent on genome rearrangements that have large effects on PFGE fingerprints (35), ST15 may be more prone to macrorestriction polymorphism than other STs of its CC (CC1) because it is the ancestor genotype.

Phenotypic biodiversity and subpopulation structure. (i) Phenotypic data set. Eighty-two variables selected as being descriptive of dairy performance, including physiological indicators (growth, acidification) and extracellular metabolic products (sugars, FAAs, organic acids, and VOCs) were assayed for the nine *L. lactis* subsp. *lactis* strains grown in pasteurized standardized cow's milk. Data for some physiological indicators (acidification and growth kinetics) were collected at various times during the milk fermentations, but most were collected after 14 days of storage (corresponding to the estimated half-life of fermented milk). Under all conditions, growth started immediately after inoculation, with maximal growth rates ranging from 0.60 to 0.80 h^{-1}. No significant difference was found for cell populations: there were 2.1×10^9 to 2.6×10^9 CFU · ml^{-1} at the end of the culture (when the pH reached 4.60). Cell cultivability after 14 days was between 50 and 70%, confirming the growth similarities of all nine strains.

The acidification properties of the nine strains were compared. As expected, the mean acidification rate of the four *L. lactis* subsp. *lactis* bv. diacetylactis strains was 40% lower than that of the other strains; this was due to the 4-fold higher acetoin concentration and lower lactic acid production of the *L. lactis* subsp. *lactis* bv. diacetylactis strains. The time taken for the fermentation to reach a pH of 4.60 was 2 h longer for the *L. lactis* subsp. *lactis* bv. diacetylactis strains than for the other strains, a delay considered to be significant for the dairy industry. Supernatant from 14-day fermented milk was assayed for 18 FAAs; all were detected, and their concentrations could be quantified. However, the biogenic amine γ-aminobutyric acid (GABA) was detected in only three strains, EIP33A (0.1 mM), DIA-A, and MC70 (1 mM). A total of 47 different VOCs, belonging to the hydrocarbon, alcohol, aldehyde, ketone, ester, sulfide compound, and free fatty acid (FFA) families, were identified in the samples and quantified. To investigate the relationships between strains and the phenotypic variables of fermented milk, an ANOVA of the 82 variables was performed. This analysis revealed that only VOCs were strongly dependent on the strains (P value of <0.05 for the ANOVA test).

(ii) Metabolic exploration of fermented milk by correlation investigations. The analysis of these 82 scaled variables as a single integrated phenotypic data set allowed exploration of the metabolic network in fermented milk. Spearman's rank-order correlations after P value multiple-testing corrections by the Benjamini-Hochberg method (P value cutoff, 0.05) were then performed to investigate pairwise associations between the variables involved in strain signatures in milk (see Fig. S2 in the supplemental material). The principal finding was a strong correlation ($R \geq 0.70$) among VOCs of the FFAs (C_4 to $C_{9:2}$), methyl-ketone (2-undecanone, 2-tridecanone, 2-pentadecanone), aldehyde (hexadecanal), alcohol (heptanol, octanol, nonanol), and lactone (δ-octalactone, δ-decalactone, δ-dodecalactone) families (see Fig. S2A). All of these metabolites are involved in the four main steps of lipolysis, i.e., release of FFAs from milk triglycerides, followed by β-oxidation of released FFAs to α-keto acids, then decarboxylation to alkan-2-ones, and finally reversible reduction to the correspond-

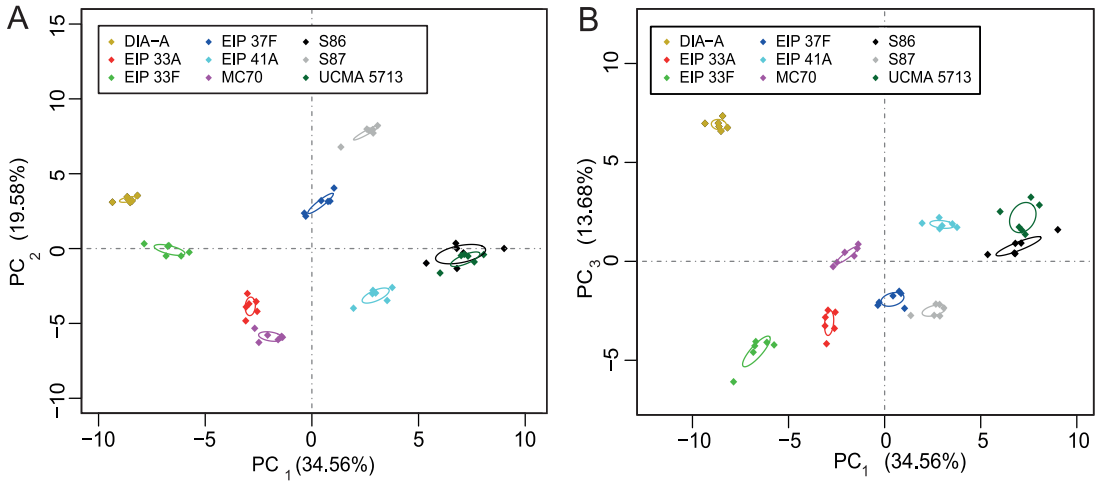

FIG 2 PCA of domesticated *L. lactis* subsp. *lactis* strains in milk from an integrative phenotypic data set. Both panels A and B are score plots of PC_1, PC_2, and PC_3 scores. Diamonds of the same color correspond to biological replicates of the same strain. Ellipses show the strain category models at a 95% confidence level.

ing alkan-2-ols (13, 36–38). The strong relationships among these metabolites confirmed that they are involved in the same pathway.

The metabolites derived from glycolysis, more specifically, those involved in pyruvate metabolism (acetaldehyde, lactic acid, acetoin, diacetyl, and butanediol), displayed strong positive correlations with each other ($R \geq 0.67$). The contribution of citrate consumption to the production of diacetyl, acetoin, acetic acid, lactic acid, and butanediol was confirmed by strong negative correlations ($R \leq -0.75$) (see Fig. S2B).

Several correlations related to proteolysis were also identified. First, the catabolic link between benzaldehyde and its amino acid precursor, phenylalanine, was confirmed by a strong negative correlation ($R < -0.70$). Metabolites of methionine catabolism (sulfide and thiol VOCs) were positively correlated (see Fig. S2B). Lastly, 3-methylbutanal and 2-methylpropanol, the main products of branched-chain amino acid catabolism, were positively correlated ($R = 0.61$).

(iii) High strain phenotypic diversity. A PCA of the 54 fermented milk samples (six cultures of each of the nine strains) was used to study phenotypic strain diversity. This analysis was limited to the first three PCs of the PCAs because the cumulative variance analysis of PC_1 to PC_3 accounted for 67.8%, a percentage sufficiently high to ensure that the PCA plots were representative of the main features of the data set. On the basis of visual explorations of the score plots (Fig. 2A and B), nine independent signatures, i.e., one per strain, were observed. This result illustrates the high phenotypic diversity of genetically closely related strains, all exhibiting a dairy phenotype (efficient growth in milk associated with fast coagulation). This result also confirms the suitability of the 82 monitored variables for establishing a discriminatory phenotypic data set for investigation of the diversity of the "domesticated" *L. lactis* subpopulation.

Thus, each of these nine strains has a unique phenotypic signature. We classified them according to their metabolic proximity. Although the PCA score plots allowed intuitive strain grouping, a clustering analysis was preferred to allow the robustness of the strain classification to be assessed. The 82 variables were used to build a dendrogram by hierarchical clustering (Fig. 3). The accuracy of the classification was checked by bootstrap analysis, which revealed that all of the biological replicates of a strain (six independent experiments) were repeatedly grouped together (bootstrap value >0.94), confirming the quality of the data. Thus, our integrated phenotypic approach allows robust and accurate strain identification and discrimination by analysis of fermented milk. It would be interesting to analyze how these nine strains could be grouped together on the basis of the proximity of their phenotypes. To choose the relevant number of strain clusters in this classification, two criteria were examined: the decrease in interclass variance (data not shown) and bootstrap analysis (Fig. 3). However, with these criteria, four, five, and nine clusters were possible options, without its being possible to conclude which was the most appropriate. Consequently, further investigations were required to determine the number of clusters and, by extension, the organization of the "domesticated" *L. lactis* subsp. *lactis* subpopulation.

(iv) Subpopulation organization and variable selection. To identify the relevant number of strain clusters in the "domesticated" *L. lactis* subsp. *lactis* subpopulation, a robustness analysis was carried out. This strain cluster analysis also aimed to reduce the number of variables used from 82, since strain discrimination is no longer required. This approach allowed the identification of the most significant phenotypic markers. We focused our study on the variables making the largest contribution, i.e., those most responsible for strain positioning. From the variable selection (for details, see Materials and Methods), 20 variables, corresponding to 25% of the 82 initial variables, was found to be the minimal number of most contributory variables allowing robust strain clustering in each case (data not shown). Four strain clusters were thereby defined (bootstrap value >0.82) and named I to IV. Their compositions are shown in a dendrogram in Fig. 3. A supplementary PLS-DA was then used to estimate the error rate of strain affiliation with the four clusters identified. With only the 25% of the variables that contributed the most to this clustering, perfect accuracy (100%) in strain cluster affiliation was predicted.

Thus, 20 of the 82 variables studied were identified as the most contributory. Variable selection through sparse PCA confirmed the suitability of the set of discriminatory variables. All of these selective variables were related to VOCs and were classified into

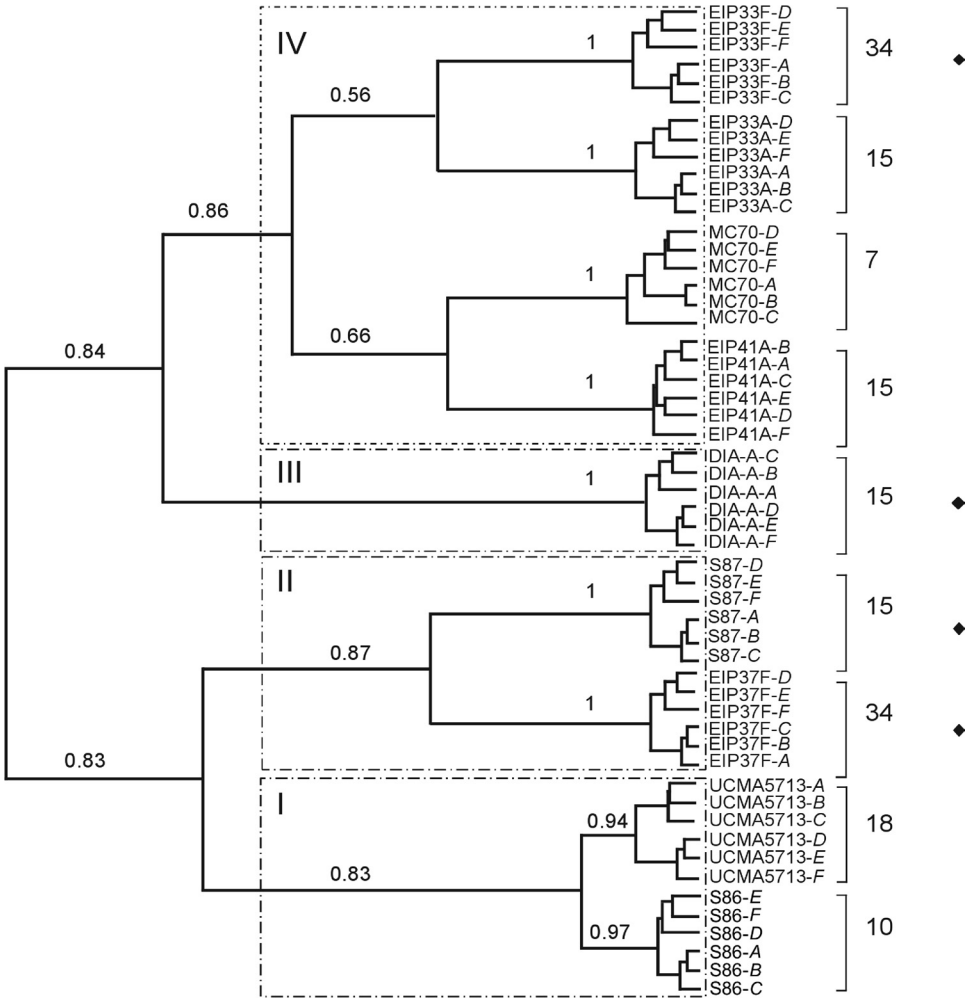

FIG 3 Integrative phenotypic classification of nine domesticated *L. lactis* subsp. *lactis* strains grown in milk. The Euclidian distance metric and Ward's criterion were used for hierarchical ascendant clustering based on 82 variables. From left to right, a strain dendrogram, the names of the 54 experiments (nine strains with their six replicates, from A to F), STs from MLST, and the presence of the *citP* gene (black diamond) are shown. Bootstrap values above 0.50 are indicated at the corresponding nodes ($n = 500$). Dotted squares with roman numerals I to IV in the left panel correspond to the four main conserved clusters identified by robustness analysis.

three main groups (A, B, and C) (Table 2). These groups are representative of the three major catabolic pathways involved in flavor formation, consistent with the findings of the correlation analysis. Group A is related to glycolysis (and citrate products except for nonan-2-one), group B is involved in proteolysis, and group C is involved in lipolysis. Analysis of the correlations between these three groups of selected variables and the first three PCs of the PCA (Table 2) demonstrated the relevance of the PCs for the analysis of metabolism.

To interpret the metabolic significance of each PC, the major coefficients of correlation between the three variable groups and each PC were investigated (Table 2). The first PC, with which all of the group C variables (plus nonan-2-one) were positively correlated, could be interpreted as the lipolytic activity of the strains in milk. Group A variables were positively correlated with the second PC, suggesting that PC2 is related to glycolysis. The third PC, correlated only with group B variables, may be interpreted as an indicator of proteolytic activity. From these results, the metabolic orientations of the signatures of single strains could be easily identified. Thus, strains UCMA5713 and S86 display the highest lipolytic activity; S87 produces the largest amounts of the end products of the citrate pathway; and strain DIA-A displays the highest concentrations of VOCs from amino acid catabolism (Fig. 2A and Table 2).

The biological significance of these four clusters was investigated according to the metabolic interpretation of each of the PCs (as described by using the 20 most contributory variables, groups A, B, and C) and by positioning the four strain clusters in PCA score plots (Fig. 2A and B). In cluster I, strains UCMA5713 and S86 produced predominantly VOCs by intense lipolysis (group C) whereas the levels of these compounds were relatively low in fermentations with cluster III strains (DIA-A). In cluster II, strains EIP37F and S87 were characterized mainly by the production of acetoin, diacetyl, and butanediol (group A). Note that the DIA-A strain was a singleton (cluster III), emphasizing its extreme phenotypic specificity. DIA-A was also associated with strong amino acid catabolism (group B). Finally, the four remaining strains (EIP41A, MC70, EIP33A, and EIP33F) belonged to cluster IV;

TABLE 2 Coefficients of correlation of the 20 most contributory variables in the PCA to the three PCs[a]

Group and variable	PC_1		PC_2		PC_3	
	R	*P* value	*R*	*P* value	*R*	*P* value
A						
Acetoin	−0.55	1.45×10^{-5}	0.77	1.54×10^{-11}		
Acetaldehyde			0.86	5.51×10^{-16}	−0.37	5.37×10^{-3}
2-Nonanone	0.31	2.19×10^{-2}	0.89	9.87×10^{-20}		
Butan-2,3-diol D^+	−0.46	5.24×10^{-4}	0.81	8.77×10^{-14}		
B						
2-Methylbutanal	−0.57	5.50×10^{-6}	0.29	3.05×10^{-2}	0.74	1.95×10^{-10}
Butyl acetate	−0.54	2.24×10^{-5}	−0.29	3.05×10^{-2}	0.76	3.83×10^{-11}
2-Methylpropan-1-ol	−0.33	1.54×10^{-2}	0.51	7.56×10^{-5}	0.73	2.64×10^{-10}
2-Hydroxypentan-3-one	−0.61	8.44×10^{-7}	0.43	1.10×10^{-5}	0.60	1.39×10^{-6}
Ethanol	−0.28	4.22×10^{-2}	−0.43	1.19×10^{-3}	0.90	8.02×10^{-21}
C						
Octanol	0.79	1.28×10^{-12}	0.53	3.21×10^{-5}		
Undecane-2-one	0.90	1.94×10^{-20}	0.36	7.82×10^{-20}		
Butanoic acid	0.96	8.93×10^{-30}				
2-Tridecanone	0.88	8.68×10^{-19}	0.44	7.48×10^{-4}		
Hexanoic acid	0.93	5.21×10^{-24}	0.28	3.73×10^{-2}		
Heptanoic acid	0.92	3.38×10^{-23}	0.35	8.78×10^{-3}		
2-Pentadecanone	0.89	8.43×10^{-20}	0.42	1.43×10^{-3}		
Octanoic acid	0.92	3.71×10^{-22}	0.34	1.23×10^{-2}		
Nonanoic acid	0.89	8.73×10^{-20}	0.37	5.51×10^{-3}		
Decanoic acid	0.93	1.00×10^{-23}	0.28	3.94×10^{-2}		
δ-Dodecalactone	0.90	2.95×10^{-29}	0.32	1.73×10^{-2}		

[a] Significance was defined as a *P* value below 0.05. *R*, coefficient of correlation of the variable to the PC. Hierarchical clustering of ascendant classification using the Euclidian distance metric and Ward's criterion is shown for groups A to C.

their phenotypic signatures are not characteristic of a particular metabolic pathway, with near mean concentrations of the most contributory VOCs. Therefore, we considered cluster IV to be characterized by no dedicated metabolic pathway.

These findings proved that within a single cluster, other phenotypic variables were responsible for strain-specific signatures. To identify the variables that are potentially specific markers of each strain, a PCA of the complete phenotypic data set for each strain cluster was performed. For metabolically undefined cluster IV, this analysis revealed that EIP41A was further specifically identified by high sulfide VOC concentrations (responsible for onion and sulfur aromas) whereas EIP33F was distinguished by high residual amino acid (His, Ala, Pro, Gln, Val, Leu, Ile, Phe) concentrations in milk. Both a high growth rate and a high ethanol concentration were characteristics of EIP33A, whereas MC70 appeared to be characterized by high furfural and 2-furanmethanol concentrations and the presence of GABA. The metabolites furfural and 2-furanmethanol can be produced in the presence of L-alanine and glucose under acidic conditions (especially at pHs ≤5) and are characterized by a "bread and burnt" aroma (39).

Genotype-phenotype integration. Phenotype-based strain clustering (Fig. 2) appeared to correspond to neither MLST phylogeny (strains of the same ST did not cluster together) nor genomic PFGE profiles (four clusters formed rather than the three groups of related strains suggested by PFGE) (Fig. 1). Moreover, the four strains belonging to *L. lactis* subsp. *lactis* bv. diacetylactis were distributed among different phenotypic clusters (II, III, and IV, Fig. 3). Comparisons of the integrated phenotypic classification (Fig. 3) with genotypic strain diversity (Fig. 1) confirmed the discrepancies among the classifications. In particular, the specificity of strain MC70 is underlined in the genotypic classification because of its specific CC2 affiliation of the "domesticated" strains, but its phenotypic signature was identified as not dedicated to a particular metabolic pathway. These results highlight the need for a system of classification integrating both genotypes and phenotypes to describe *L. lactis* subsp. *lactis* biodiversity.

Using the integrated genotype-phenotype data set, strains were subjected to both PCA and clustering analysis (Fig. 4). This led to a five-cluster organization rather than the four strain clusters (I to IV) generated with discriminatory phenotypic markers. Two groups in these two classifications were similar (clusters II and III). The previous cluster IV was the most affected by the new classification. In particular, MC70, which was previously classified in cluster IV, was now singled out. The variable analysis underlined the contribution of MLST phylogenic distances to PC2 (data not shown) to this new classification and thus the position of MC70. The use of independent genetic (MLST) and genomic (PFGE) data sets in addition to phenotypic variables revealed the relevance of MLST in the classification. Indeed, unlike MLST data, the PFGE genomic diversity data set did not change the phenotypic PCA or clustering results (data not shown). The other strains previously classified in cluster IV (EIP41A and EIP33F) were classified in existing clusters (I and IV) by the integrated genotype-phenotype classification.

Variable selection (as previously described) demonstrated that no genotypic variable (either MLST or PFGE distances) was selected as being one of the most contributory variables. The con-

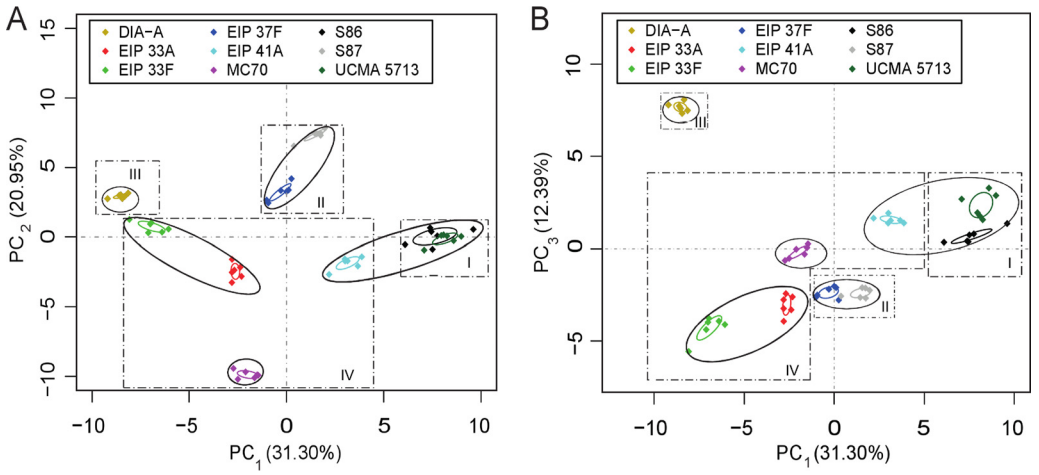

FIG 4 Integrative phenotypic and genotypic PCA of nine domesticated *L. lactis* subsp. *lactis* strains grown in milk. Score plots of PC_1-PC_2 (A) and PC_1-PC_3 (B) are shown. Diamonds with the same color are biological replicates of the same strain, and colored ellipses show the strain category models (at a 95% confidence level). Dotted squares with roman numerals I to IV correspond to the four phenotypic clusters (Fig. 3), and black ellipses represent the five integrative genotype-phenotype clusters.

tributory variables were, again, mostly VOCs (21/23), acetic acid, and fermentation time.

DISCUSSION

We report an analysis of the phenotypic diversity of nine strains of *L. lactis* subsp. *lactis* representative of the "domesticated" ecotype and exhibiting a dairy phenotype (efficient growth in milk and fast coagulation). Eighty-two variables related to important dairy features, including physiological descriptors (growth, acidification) (10), carbon and nitrogen metabolites, and flavor determinants (14, 40), were monitored in 54 fermentations (6 per strain). Our integrated phenotypic approach confirms the suitability of these 82 variables for the establishment of well-characterized dairy strain signatures in fermented milk. Furthermore, this study reveals large strain-to-strain phenotypic divergences that are sufficient for the accurate identification of the nine strains from their milk fermentation signatures. Consequently, these 82 variables provide a useful data set for the classification of dairy strains and provide new possibilities for applications, particularly for strain identification and selection.

Unexpectedly, these results illustrate the substantial phenotypic diversity in a limited genetic subpopulation of strains. Indeed, phylogenetic studies suggest that dairy strains are poor contributors to the overall genetic diversity of *L. lactis* subsp. *lactis* (8, 21, 41). "Domesticated" strains arose from a genetic bottleneck or were the result of the successful adaptation to milk of a common "environmental" ancestor due to genomic flexibility (6, 42, 43). Consequently, it has been suggested that the phenotypic diversity of dairy strains is expected to be low because of the small pool of strains (44, 45). However, despite the extensive genotypic similarities, we report substantial phenotypic divergence between "domesticated" strains: consequently, these strains have different potentials for different industrial applications.

The phenotypic biodiversity of these dairy strains could be subdivided into four robust clusters based on metabolic orientations (lipolysis, proteolysis, and glycolysis). These clusters may be helpful in assessing strain relatedness. The analysis of more *L. lactis* subsp. *lactis* collections, including strains representative of the "domesticated" ecotype from several dairy environments, would probably lead to the compositions of these clusters being modified and to the identification of additional clusters.

However, these findings imply that it would be useful to have appropriate markers to define the specific abilities of a strain. Indeed, a clear description of the differences between strain groups is necessary for better control of starter choice. Therefore, we used the integrated phenotypic data set to identify the 20 variables that contribute the most to discrimination between the strains. All of these variables are related to VOCs. Our study confirms that selected VOCs can be used as discriminatory phenotypic markers to predict strain signatures in milk processing (46). Combined with the recent development of ultrafast GC-time of flight MS technology for fermented milk (46, 47), monitoring of these markers should help in the high-throughput screening of strain collections. Further analyses of these 20 VOCs revealed that the three major pathways involved in flavor, i.e., lipolysis, proteolysis, and glycolysis, were represented. Proteolysis and glycolysis in milk have been studied and described (13, 48, 49), whereas lipolysis in fermented milk has been regularly underestimated; this is largely because many experiments are carried out with skim milk media. This work underlines the importance of lipolysis, because more than half of the most discriminatory VOC markers belonged to this class.

L. lactis subsp. *lactis* bv. diacetylactis strains are used in the dairy industry to produce acetoin and notably diacetyl, which imparts a high level of buttery flavor notes. Surprisingly, the four strains in our sample belonging to *L. lactis* subsp. *lactis* bv. diacetylactis (DIA-A, EIP33F, S87, and EIP33F) were not grouped in a single cluster at a phenotypic level but were classified together by integrated genotypic classification. Interestingly, this feature demonstrates that acetoin and diacetyl production is not relevant for discriminating *L. lactis* subsp. *lactis* bv. diacetylactis strain signatures in milk at a phenotypic level. A recent study of diacetyl- and acetoin-producing *L. lactis* strains isolated from diverse origins indicated that the production of this aroma is not restricted to strains able to grow in milk (50). These observations illustrate the intragroup metabolic diversity of these strains and show the need

for further metabolic investigations to characterize these strains thoroughly (notably, amino acid and lipid catabolism).

The phenotypic classification was not entirely consistent with the genetic (MLST) and genomic (PFGE) dendrograms or with the new integrated genotypic classification. The proposed integrated genotypic classification of these dairy strains was coherent with the CC phylogenic organization of "domesticated" strains (MLST) (6), discrimination of the *L. lactis* subsp. *lactis* bv. diacetylactis lineage (based on results of PCR amplification of *citP*), and strain genomic uniqueness (nonredundancy of SmaI macrorestriction by PFGE).

Genotype and phenotype matching is not straightforward, as previously demonstrated, because of the regulation of genome and proteome expression (18, 21, 43). However, we identified contributory phenotypic determinants that are promising for the discovery of reliable genetic markers. This approach has been used previously but only at the *L. lactis* subsp. *lactis* and *cremoris* level (9, 22). On the basis of the three groups of VOCs identified as the most contributory variables, the corresponding genes (i.e., those involved in lipolysis, proteolysis, and glycolysis and their regulation) could be sequenced and their variability analyzed.

We also considered a genotype-phenotype classification to integrate all aspects of dairy *L. lactis* subsp. *lactis* diversity. We found that the genotype made only a small contribution to diversity, which depends on large phenotypic differences. Nevertheless, integration of genotypic diversity modified the phenotypic classification and led to a five-cluster organization rather than the four strain clusters generated with only discriminatory phenotypic markers. In particular, this increased the diversity of the strains "not dedicated to a metabolic pathway" of cluster IV. Moreover, "domesticated" strains of L. *lactis* subsp. *lactis* are characterized by noticeable mobility and substantial genomic variability of their genomes (6). This genotypic variability may contribute significantly to variations in the properties of the starter culture. Adding genotypic characteristics would ensure that all of the strain potentialities encoded in the genotype would be covered, including those not expressed in our model dairy process but which may be of relevance to other technological applications (such as osmotic resistance in cheese ripening or phage resistance). The genotypic-phenotypic classification tends toward a predictive classification for starter manufacturing.

Finally, this study illustrates the feasibility of establishing well-characterized strain signatures. This approach could be extended to mixed-culture profiling. Indeed, it may be possible to assess the effect of strain adjunction in cocultures of lactococci by comparing single-strain signatures to the resulting mix signatures. This method may complement traditional trial-and-error methods used to design targeted starters for particular applications and facilitate the rational development of defined mixed cultures.

ACKNOWLEDGMENTS

We acknowledge Helene Tormo from INP-EI Purpan (Toulouse, France) for providing EIP bacterial strains. We thank Gilles De Revel and Laurent Riquier from the Analytical Chemical Service of ISVV for SPME-GC-MS analyses.

This research was supported by Yoplait S.A.

REFERENCES

1. **Schleifer KH, Kraus J, Dvorak C, Kilpper-Bälz R, Collins MD, Fischer W.** 1985. Transfer of *Streptococcus lactis* and related streptococci to the genus *Lactococcus* gen. nov. Syst. Appl. Microbiol. **6:**183–195.
2. **Klijn N, Weerkamp AH, de Vos WM.** 1995. Detection and characterization of lactose-utilizing *Lactococcus* spp. in natural ecosystems. Appl. Environ. Microbiol. **61:**788–792.
3. **Kelly W, Ward L.** 2002. Genotypic vs. phenotypic biodiversity in *Lactococcus lactis*. Microbiology **148:**3332–3333.
4. **Salama MS, Sandine WE, Giovannoni SJ.** 1993. Isolation of *Lactococcus lactis* subsp. *cremoris* from nature by colony hybridization with rRNA probes. Appl. Environ. Microbiol. **59:**3941–3945.
5. **Rademaker JL, Herbet H, Starrenburg MJ, Naser SM, Gevers D, Kelly WJ, Hugenholtz J, Swings J, van Hylckama Vlieg JE.** 2007. Diversity analysis of dairy and nondairy *Lactococcus lactis* isolates, using a novel multilocus sequence analysis scheme and $(GTG)_5$-PCR fingerprinting. Appl. Environ. Microbiol. **73:**7128–7137.
6. **Passerini D, Beltramo C, Coddeville M, Quentin Y, Ritzenthaler P, Daveran-Mingot ML, Le Bourgeois P.** 2010. Genes but not genomes reveal bacterial domestication of *Lactococcus lactis*. PLoS One **5:**e15306. doi:10.1371/journal.pone.0015306.
7. **Fernández E, Alegría A, Delgado S, Martín MC, Mayo B.** 2011. Comparative phenotypic and molecular genetic profiling of wild *Lactococcus lactis* subsp. *lactis* strains of the *L. lactis* subsp. *lactis* and *L. lactis* subsp. *cremoris* genotypes, isolated from starter-free cheeses made of raw milk. Appl. Environ. Microbiol. **77:**5324–5335.
8. **Siezen RJ, Bayjanov JR, Felis GE, van der Sijde MR, Starrenburg M, Molenaar D, Wels M, van Hijum SA, van Hylckama Vlieg JE.** 2011. Genome-scale diversity and niche adaptation analysis of *Lactococcus lactis* by comparative genome hybridization using multi-strain arrays. Microb. Biotechnol. **4:**383–402.
9. **Taïbi A, Dabour N, Lamoureux M, Roy D, LaPointe G.** 2010. Evaluation of the genetic polymorphism among *Lactococcus lactis* subsp. *cremoris* strains using comparative genomic hybridization and multilocus sequence analysis. Int. J. Food Microbiol. **144:**20–28.
10. **Cachon R, Jeanson S, Aldarf M, Divies C.** 2002. Characterisation of lactic starters based on acidification and reduction activities. Lait **82:**281–288.
11. **Brasca M, Morandi S, Lodi R, Tamburini A.** 2007. Redox potential to discriminate among species of lactic acid bacteria. J. Appl. Microbiol. **103:**1516–1524.
12. **Kristo E, Biliaderis CG, Tzanetakis N.** 2003. Modelling of the acidification process and rheological properties of milk fermented with a yogurt starter culture using response surface methodology. Food Chem. **83:**437–446.
13. **Cheng H.** 2010. Volatile flavor compounds in yogurt: a review. Crit. Rev. Food Sci. Nutr. **50:**938–950.
14. **Smit G, Smit BA, Engels WJ.** 2005. Flavour formation by lactic acid bacteria and biochemical flavour profiling of cheese products. FEMS Microbiol. Rev. **29:**591–610.
15. **Aly S, Floury J, Piot M, Lortal S, Jeanson S.** 2012. The efficacy of nisin can drastically vary when produced in situ in model cheeses. Food Microbiol. **32:**185–190.
16. **Hickey RM, Ross RP, Hill C.** 2004. Controlled autolysis and enzyme release in a recombinant lactococcal strain expressing the metalloendopeptidase enterolysin A. Appl. Environ. Microbiol. **70:**1744–1748.
17. **Valence F, Deutsch SM, Richoux R, Gagnaire V, Lortal S.** 2000. Autolysis and related proteolysis in Swiss cheese for two Lactobacillus helveticus strains. J. Dairy Res. **67:**261–271.
18. **Bachmann H, Starrenburg MJ, Dijkstra A, Molenaar D, Kleerebezem M, Rademaker JL, van Hylckama Vlieg JE.** 2009. Regulatory phenotyping reveals important diversity within the species *Lactococcus lactis*. Appl. Environ. Microbiol. **75:**5687–5694.
19. **Bayjanov JR, Molenaar D, Tzeneva V, Siezen RJ, van Hijum SA.** 2012. PhenoLink—a web-tool for linking phenotype to ~omics data for bacteria: application to gene-trait matching for *Lactobacillus plantarum* strains. BMC Genomics **13:**170. doi:10.1186/1471-2164-13-170.
20. **Bayjanov JR, Starrenburg MJ, van der Sijde MR, Siezen RJ, van Hijum SA.** 2013. Genotype-phenotype matching analysis of 38 *Lactococcus lactis* strains using random forest methods. BMC Microbiol. **13:**68. doi:10.1186/1471-2180-13-68.
21. **Tan-a-ram P, Cardoso T, Daveran-Mingot ML, Kanchanatawee S, Loubière P, Girbal L, Cocaign-Bousquet M.** 2011. Assessment of the diversity of dairy *Lactococcus lactis* subsp. *lactis* isolates by an integrated approach combining phenotypic, genomic, and transcriptomic analyses. Appl. Environ. Microbiol. **77:**739–748.
22. **Kütahya OE, Starrenburg MJ, Rademaker JL, Klaassen CH, van Hylckama Vlieg JE, Smid EJ, Kleerebezem M.** 2011. High-resolution ampli-

fied fragment length polymorphism typing of *Lactococcus lactis* strains enables identification of genetic markers for subspecies-related phenotypes. Appl. Environ. Microbiol. **77:**5192–5198.
23. **Taïbi A, Dabour N, Lamoureux M, Roy D, LaPointe G.** 2011. Comparative transcriptome analysis of Lactococcus lactis subsp. cremoris strains under conditions simulating cheddar cheese manufacture. Int. J. Food Microbiol. **146:**263–275.
24. **Cocaign-Bousquet M, Lindley ND.** 1995. Pyruvate overflow and carbon flux within the central metabolic pathways of *Corynebacterium glutamicum* during growth on lactate. Enzyme Microb. Technol. **17:**260–267.
25. **Nouaille S, Even S, Charlier C, Le Loir Y, Cocaign-Bousquet M, Loubière P.** 2009. Transcriptomic response of *Lactococcus lactis* in mixed culture with *Staphylococcus aureus*. Appl. Environ. Microbiol. **75:**4473–4482.
26. **Tanigawa K, Kawabata H, Watanabe K.** 2010. Identification and typing of *Lactococcus lactis* by matrix-assisted laser desorption ionization–time of flight mass spectrometry. Appl. Environ. Microbiol. **76:**4055–4062.
27. **Le Bourgeois P, Lautier M, van den Berghe L, Gasson MJ, Ritzenthaler P.** 1995. Physical and genetic map of the *Lactococcus lactis* subsp. *cremoris* MG1363 chromosome: comparison with that of *Lactococcus lactis* subsp. *lactis* IL 1403 reveals a large genome inversion. J. Bacteriol. **177:**2840–2850.
28. **Nelson RJ.** 1994. Relationship between phylogeny and pathotype for the bacterial blight pathogen of rice. Appl. Environ. Microbiol. **60:**3275–3283.
29. **Felsenstein J.** 1989. PHYLIP—phylogeny inference package (version 3.2). Cladistics **5:**164–166.
30. **Benjamini Y, Hochberg Y.** 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol. **57:**289–300.
31. **Lê Cao K-A, González I, Déjean S.** 2009. integrOmics: an R package to unravel relationships between two omics datasets. Bioinformatics **25:**2855–2856.
32. **Ward JH.** 1963. Hierarchical grouping to optimize an objective function. J. Am. Stat. Assoc. **58:**236–244.
33. **Shen H, Huang JZ.** 2008. Sparse principal component analysis via regularized low rank matrix approximation. J. Multivariate Anal. **99:**1015–1034.
34. **David S, van der Rest ME, Driessen AJ, Simons G, de Vos WM.** 1990. Nucleotide sequence and expression in *Escherichia coli* of the *Lactococcus lactis* citrate permease gene. J. Bacteriol. **172:**5789–5794.
35. **Le Bourgeois P, Daveran-Mingot ML, Ritzenthaler P.** 2000. Genome plasticity among related *Lactococcus* strains: identification of genetic events associated with macrorestriction polymorphisms. J. Bacteriol. **182:**2481–2491.
36. **Collins YF, McSweeney PL, Wilkinson MG.** 2003. Lipolysis and free fatty acid catabolism in cheese: a review of current knowledge. Int. Dairy J. **13:**841–866.
37. **Fox PF, Wallace JM.** 1997. Formation of flavor compounds in cheese. Adv. Appl. Microbiol. **45:**17–85.
38. **McSweeney PL, Sousa MJ.** 2000. Biochemical pathways for the production of flavour compounds in cheeses during ripening: a review. Lait **80:**32.
39. **Bemiller J, Huber K.** 2007. Carbohydrates, p 83–155. *In* Damodaran S, Parkin KL, Fennema OR (ed), Fennema's food chemistry, 4th ed. CRC Press, New York, NY.
40. **Gutiérrez-Méndez N, Vallejo-Cordoba B, González-Córdova AF, Nevárez-Moorillón GV, Rivera-Chavira B.** 2008. Evaluation of aroma generation of *Lactococcus lactis* with an electronic nose and sensory analysis. J. Dairy Sci. **91:**49–57.
41. **Bayjanov JR, Wels M, Starrenburg M, van Hylckama Vlieg JE, Siezen RJ, Molenaar D.** 2009. PanCGH: a genotype-calling algorithm for pan-genome CGH data. Bioinformatics **25:**309–314.
42. **van Hylckama Vlieg JE, Rademaker JL, Bachmann H, Molenaar D, Kelly WJ, Siezen RJ.** 2006. Natural diversity and adaptive responses of *Lactococcus lactis*. Curr. Opin. Biotechnol. **17:**183–190.
43. **Kelly WJ, Ward LJ, Leahy SC.** 2010. Chromosomal diversity in *Lactococcus lactis* and the origin of dairy starter cultures. Genome Biol. Evol. **2:**729–744.
44. **Salama MS, Musafija-Jeknic T, Sandine WE, Giovannoni SJ.** 1995. An ecological study of lactic acid bacteria: isolation of new strains of *Lactococcus* including *Lactococcus lactis* subspecies *cremoris*. J. Dairy Sci. **78:**1004–1017.
45. **Hansen EB.** 2002. Commercial bacterial starter cultures for fermented foods of the future. Int. J. Food Microbiol. **78:**119–131.
46. **de Bok FA, Janssen PW, Bayjanov JR, Sieuwerts S, Lommen A, van Hylckama Vlieg JE, Molenaar D.** 2011. Volatile compound fingerprinting of mixed-culture fermentations. Appl. Environ. Microbiol. **77:**6233–6239.
47. **Smit BA, Engels WJ, Bruinsma J, van Hylckama Vlieg JE, Wouters JT, Smit G.** 2004. Development of a high throughput screening method to test flavour-forming capabilities of anaerobic micro-organisms. J. Appl. Microbiol. **97:**306–313.
48. **Chipman D, Barak Z, Schloss JV.** 1998. Biosynthesis of 2-aceto-2-hydroxy acids: acetolactate synthases and acetohydroxyacid synthases. Biochim. Biophys. Acta **1385:**401–419.
49. **Smit BA, Engels WJ, Smit G.** 2009. Branched chain aldehydes: production and breakdown pathways and relevance for flavour in foods. Appl. Microbiol. Biotechnol. **81:**987–999.
50. **Passerini D, Laroute V, Coddeville M, Le Bourgeois P, Loubière P, Ritzenthaler P, Cocaign-Bousquet M, Daveran-Mingot ML.** 2013. New insights into Lactococcus lactis diacetyl- and acetoin-producing strains isolated from diverse origins. Int. J. Food Microbiol. **160:**329–336.